DOCUMENT RESUME

ED 201 200 FL 012 230

AUTHOR Dubois, David D.
TITLE The Children's English and Services Study: A Methodological Review.
INSTITUTION National Center for Education Statistics (DHEW), Washington, D.C.
PUB DATE Aug 80
NOTE 54p.; For related documents, see ED 193 971-972. Some faint type.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS English (Second Language); *Evaluation Methods; Language Proficiency; *Testing
IDENTIFIERS *Childrens English and Services Study; Limited English Speaking

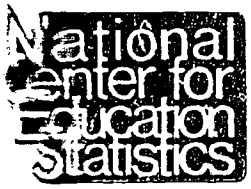
ABSTRACT

The Children's English and Services Study was a project designed to assess the bilingual education needs of limited English speaking children in the United States. The submission of a draft final report prompted the present report from the sponsoring organization, in which various methodological procedures are questioned and recommendations are made for the revision of the final report. The three analytical issues involved are: (1) were the items selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly? (2) were the cutoff scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency (LEP) set properly? and (3) what were the effects of non-response bias on the counts and estimates of the number of LEP children? With regard to (1), it is recommended that certain caveats be set forth in the final report. Criteria are introduced that, with respect to (2) actually revise the figures regarding the number of LEP children. Further investigations of nonresponse bias were found not to be warranted. (JB)

The Children's English and Services Study: A Methodological Review

by

Dr. David D. Dubois
Education Policy Fellow



August 1980

CONTENTS

INTRODUCTION

On January 21, 1980, the Office of Research and Analysis (ORA) of the National Center for Education Statistics (NCES), U.S. Department of Education, issued a report entitled Analytical Issues Regarding the Children's English and Services Study (AI/CESS). The purpose of the AI/CESS report was:

> . . . to present and discuss three analytical issues which have been identified as a result of a post hoc assessment of the research design, data analyses and other information which are described in the 1978 Children's English and Services Study (CESS) Draft Report of September 6, 1979 (and a later revision dated November 1979).[1]

A copy of the AI/CESS report is found in appendix A.

The objective of the present NCES/ORA inquiry is best summarized by the following passage from the January report:

> Since the results of the 1978 CESS are of tremendous importance to present and future research studies, bilingual (education) program and policy development, and funding for bilingual education, unresolved analytical issues which could adversely affect the validity of the results are being stated with the hope of their resolution.[2]

It is generally recognized that secondary analyses of data and research designs frequently reveal analysis errors or areas of skepticism in the design. Sterling and Weinkam (1979), who discovered misclassifications in a study of mortality among U. S. veterans, describe the potential response of managers to this discovery as either "cooperative" or "adversary."

In the former case, an attempt is made to determine the source of concern and to restructure the procedures or analyses. In the latter case, attempts are made to eliminate the discovery of errors rather than their source.

Regarding this problem, Sterling and Weinkam further observed that:

> ... there may be underlying sociological and psychological forces operating which make it more acceptable for management to adopt an adversary rather than a cooperative stance even in scientific instances. From a simpleminded perspective, to acknowledge the existence of errors may require considerable effort and expenditures to correct them, not to say anything of extracting accountability from some individuals who insist on bringing these errors to the attention of management as troublemakers.[3]

They continued by indicating that "as the value of [illegible] data files for secondary analysis become increasingly clear [illegible] a great deal of value in the use of properly collected and suitable [illegible] aspect that other discoveries similar to ours will be made." They recommended that a mechanism be established for the encouragement of [illegible] analysis for the validation of the appropriateness of preceding work [illegible] tribution of the results of these analyses.

NCES/ORA recognizes that many research reports are [illegible] inaccurate in one or more ways and the CESS Draft Report [illegible] no exception. Accordingly, NCES ORA's purpose on this report is [illegible] issues and make specific recommendations for modifications [illegible] report.

This report includes a discussion of the three [illegible] in the AI/CESS report and will be based, in part, upon the responses received from it. In addition, pertinent literature on language development [illegible] assessment, the relationship of language acquisition to cognition, [illegible] nal data analyses completed by NCES provide the bases for this paper

Responses to the AI/CESS report by those who were invited to respond are available for examination at the Office of Research and Analysis at the NCES. These responses will be retained on file for a period of one calendar year following the publication of this report.

STATEMENT OF THE ISSUES

The following analytical issues are the subject of this position paper:

1. Were the items which were selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?

2. Were the cutoff scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency (LEP), set properly?

3. What were the effects of measurement bias on the counts and estimates of the number of LEP children?

DISCUSSION OF THE ISSUES AND RECOMMENDATIONS

ISSUE: Were the items which were selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?

This issue was restated as:

> Is English language proficiency the dimension on which the scores vary, or are other dimensions associated with variations in the scores?[5]

Two subissues were posed, namely:

- Are the test scores related to language dominance?
- Are the test scores related to general language development?

Discussion Lourdes Miranda, President of L. Miranda and Associates, the prime contractor for the CESS, responded to the AI/CESS report. In discussing the rationale for the test items selected for the LM&AI, Miranda noted that "it was essential for us to measure the ability [of language minority children] to successfully deal with academic classroom tasks that are often as clearly reliant on memory and cognitive abilities as on English language skills."[6] Therefore, "other dimensions [e.g., cognitive] are associated with score variation."[7]

The LM&AI was specifically designed to meet the definition of limited English proficiency found in the Bilingual Education Act, that is, the 1965 Elementary and Secondary Education Act (ESEA), Section 703(a)(1)(B), as amended. The 1978 Amendment of the Act expanded the language skill domains to include speaking, reading, writing and understanding the English language. By virtue of their

limited English proficiency, Congress concluded, language minority children were denied the opportunity to attain levels of ability comparable to others at their appropriate age and grade levels.

J. Michael O'Malley, the NIE Project Officer for CESS, responded to the first issue that:

> Because functioning in the classroom often requires conceptual skills as well as oral language and literacy, the inclusion of cognitive demand in the test it was seen as an acceptable approach for increasing the content and perhaps the predictive validity of the test.[8]

O'Malley also stated that:

> A "pure" measure of English proficiency could not have possessed the content validity required to identify language minority children who had difficulty profiting from instruction in English.[9]

Earlier in his response, O'Malley said, "Simply stated, the test scores are predictive of the ability to profit from English language instruction, which determines eligibility for ESEA Title VII."[10] He observed that, "School decisions about eligibility for ESEA Title VII are often based on a child's general level of functioning in the classroom rather than on English language proficiency alone."[11] And later, "The LM&AI used tested skills in English ...to simulate the decisions schools would make in determining that language minority students could not profit from instruction in English."[12]

A review of recent literature in the areas of language assessment, linguistic and intellectual development, and bilingual education programs revealed that analytical questions in these areas have, for some time, presented serious intellectual challenges to researchers and educators. The issues raised in the AI/CESS report were presented within the framework of the CESS development process and with the knowledge that there are many unanswered basic research questions in the three areas mentioned above. NCES' purpose in this report is to clarify current thought on this issue.

The first subissue raised in the AI/CESS report was stated as: "Are the scores related to language dominance?"[13] O'Malley takes the following position:

> By exclusion in the [legislative] definition of eligibility, language dominance has no role in policy determination for ESEA Title VII eligibility. Thus, the statements in the NIE report on the CESS that language dominance was considered irrelevant is understandable.[14]

NCES/CRA believes there is justification for excluding a language dominance concept in the development of the CESS. Specifically, there does not appear to be agreement among linguists as to an operational definition and, therefore, the impact of "language dominance" upon the ability of language minority children to profit from instruction in English.

DeAvila and Duncan (1976) argue against using a "language dominance" concept when discussing school achievement. They say, "how does the concept of [language] dominance clarify the relation between the child's linguistic development and school achievement in such a way that we can do something about it?"[15] They continue by saying, "Another way of asking this question is by asking whether or not 'dominance' in or of itself determines whether what is learned or what can be learned."[16] Language dominance "does not address the real issue that the child might have language development problems in both languages--the native language and English."[17]

Some experts have argued that a language dominance concept is meaningful only when the use of a language is considered within a social or cultural context, such as: home and family relationships, social interactions, an academic domain, a business environment, or within a religious context. The degree of fluency or level of language dominance is meaningful only when the purpose for which language is being used is also stated. In this sense, several "dominance" levels might be defined.

Regarding a child's possible difficulty with both languages, Dubois (1980) states: "Whether it is appropriate to assess English language proficiency, ignoring the child's proficiency in another language remains a policy question to be addressed."[18] More specifically, this is an empirical question.

A recent article by Cummins (1979) addresses this question. In the following excerpt L_1 refers to a child's first language and L_2 refers to the second language. Cummins says:

> The lack of concern for the developmental interrelationships between language and thought in the bilingual child is one of the major reasons why evaluations and research have provided so little data on the dynamics of the bilingual child's interaction with his educational environment. A direct determinant of the quality of this interaction is clearly the level of L_1 and L_2 competence which the bilingual child develops over the course of his school career. ...What level of L_2 competence must the child possess at various grade levels in order to benefit optimally from instruction in that language? ...To what extent are L_1 and L_2 skills interdependent and what are the implications of possible interdependencies for cognitive and academic progress? In other words, do children who maintain and develop their L_1 in school develop higher or lower L_2 levels of skills than those whose L_1 is replaced by their L_2?[19]

Cummins provides research evidence for a "developmental interdependence hypothesis" which says that the level of L_2 competence which a bilingual child attains is partially a function of the type of competence the child has developed in L_1 at the time when intensive exposure to L_2 begins. In this sense, a measure of L_1 proficiency is important for policy decisions.

W. E. Lambert (1975) suggested that children exhibit either "additive bilingualism" or "subtractive bilingualism." A child's bilingualism is most likely to be "additive" when L_1 is prestigious or the "dominant" language and is, therefore, not in danger of being replaced by L_2. In this case, a bilingual child "adds" L_2 skills without a loss of L_1 skills. "Subtractive bilingualism" refers to the form of bilingualism children experience when their L_1 is eventually replaced by L_2. This is generally true when the child's L_1 is a nonprestigious or a minority language. Socioeconomic status also seems to be a factor which is related to whether a child's bilingualism is subtractive or additive. Children from upper or middle class socioeconomic strata, when given instruction in L_2, tend to experience "additive bilingualism" while children from lower socioeconomic strata tend to experience "subtractive bilingualism." Troike (1980) hypothesizes that, for children from lower socioeconomic groups, a child's cognitive development can become disrupted when a child begins learning L_2 between the ages of 6-10. Socioeconomic status and socio-political status are, therefore, related to language and cognitive development.

The second subissue stated in the AI/CESS report, was: "Are the test scores related to general language development?"[20] The concern was for the inclusion of test items on the LM&AI which included cognitive components. Miranda noted earlier that the purpose of the LM&AI was to measure the ability of language minority children to successfully deal with academic classroom language skills. In reply to this subissue, Miranda stated that, "it is difficult to imagine how a test of 'pure' linguistic competence could have been developed should we have been asked to do so."[21]

DeAvila, et al. (1979) observed that "much confusion abounds with respect to both the meaning and the measurement of English language proficiency."[22] Moreover, they noted that "the role of language and cognition in general is itself not clearly agreed upon."[23] For the purpose of this discussion, cognition shall mean the act or process of perceiving or knowing.

Cazden (1972) addressed two controversial items of interest in Child Language and Education. The first item concerns whether a person's thought is affected by the particular language forms or speech patterns with which they are familiar. The second item concerns the question of which develops first, the nonverbal idea or the words to express it.[24] Essentially, this poses the central issue: Which develops first, language or cognition? Language experts, educational psychologists and professionals in related fields apparently do not agree upon the proposed answers to this question.

Cazden's first item is based upon the Whorfian (1956) hypothesis which says that "language influences our perceptions of and responses to the world."[25] This leads us to believe that no learning can take place until language proficiency is attained; therefore, language determines cognition. Regarding Cazden's second item, Jean Piaget indicates that it is a child's cognitive

development which is the primary factor in language acquisition and development with a later emphasis on a more balanced interaction between the two.

Piaget's position is that cognition develops as a result of experience. He believes that although language contributes to further development, it is the use of language that is determined by development and not the converse.[26]

Cummins (1979), in a summary of research evidence on the role of language and cognitive development, was led to conclude:

> ... that the level of competence bilingual children achieve in their two languages acts as an intervening variable in mediating the effects of their bilingual learning experiences on cognition. Specifically, there may be threshold levels of linguistic competence which bilingual children must attain both in order to avoid cognitive deficits and to allow the potentially beneficial aspects of becoming bilingual to influence their cognitive growth.[27]

DeAvila, et al. (1979) stated that: "Edmonds (1976) has recently argued that a full understanding of language acquisition will not emerge until the process is viewed within a larger developmental framework."[28] And, related to this, "Tremaine (1975) has examined 'syntax as an instance of operational intelligence' defined in the Piagetian sense. The results indicated that children at the operational level performed significantly better in terms of syntax comprehension than children classified as nonoperational."[29] Later, DeAvila interprets Tremaine's findings as follows: "What this means is that solutions which focus on English language deficits will be of limited success as long as developmental factors are not taken into account."[30]

Studies have focused on several of these complex relationships. One of these studies (DeAvila, et al., 1979) examined the relationship between the degree of bilingualism (relative linguistic proficiency in English and Spanish), level of intellectual development (cognition), and performance on two tests of cognitive-perceptual functioning or field dependence/independence.

DeAvila concluded that: "In terms of educational implications, the most accurate and least value-laden interpretation of the findings would be to conclude that there seems to be a positive interaction between relative linguistic proficiency and cognitive/perceptual functioning."[31]

In summary, G. Richard Tucker (1979) of the Center for Applied Linguistics makes the following comments, with which the NCES/ORA agrees:

> Nor, in my opinion, have we managed to devise appropriate and valid instruments to assess language proficiency. What does it mean to know and to be able to communicate effectively and acceptably in a language? Does there exist some necessary (measureable) threshold of target language proficiency which must be attained before one is able to profit from instruction in that language? Obviously a great deal of additional interdisciplinary research is needed to examine the effects of factors such as intellectual potential, social status, physical or emotional development, age of entry, presence of native speakers, community stereotypes, teacher characteristics, classroom techniques, sequencing of languages, and social setting on the desirability and efficacy of bilingual education programs. I remain optimistic that the proposed Center for Bilingual Research may begin to move us in the right direction.[32]

Troike (1980) suggests that the effect of the density of a specific language minority group upon language proficiency in L_1 or L_2 is an additional factor to add to Tucker's list which deserves additional research attention.

NCES/ORA cannot determine the effect of the cognitive components in the LM&AI on the test scores based upon the information we now have from discussions with experts in language development and assessment, and a review of pertinent literature. A post hoc study of the cognitive component could be completed using a sample of subjects from the population which was used for the calibration of the LM&AI. This would be at an additional cost to the Government. However, the quality of the results of such a study would probably not warrant the cost since tests of language proficiency are generally confounded with language and other factors.

Recommendation NCES/ORA recommends that NIE state in the final CESS report the caveats found in our discussion of this issue. There are clearly limitations to the CESS results which are a function of the current state-of-the-art in the assessment of language proficiency.

ISSUE: Were the cutoff (critical) scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency, set properly?

Discussion The purpose of the LM&AI was to provide a mechanism for categorizing a child as being either English proficient or limited English proficient. Therefore, the critical score determined for each age-level test of the LM&AI is essential for the determination of valid LEP counts. The critical score was that score which best differentiated LEP children from fluent English-speaking (FES) children who were clearly profiting from instruction in English. As an example, if the critical score on each age-level test is lowered by two items, the estimated count of LEP children decreases from 2.41 million to 2.13 million children, or a decrease of 280,000. Similarly, if the critical score for each age-level test is raised by two items, the estimated count of LEP children is increased from 2.41 million to 2.62 million children, or an increase of 210,000. Thus, a score difference of four items has the effect of altering the count by nearly one-half million.

The NCES/ORA requested the raw data on student scores from Field Test III, which were used to determine the critical scores for the LM&AI, from the prime contractor L. Miranda and Associates, Inc. Based upon an examination of these raw data and a comparison of these findings with table A-4 of the NIE Draft Report on the CESS, a discrepancy in the data of table A-4 was discovered. This discrepancy was called to the attention of the prime contractor. Ms. Miranda replied that the procedure used for determining the critical scores, based on a discriminant function analysis, was a modification (Grand Mean - Constant = Cutoff) of the more conventional approach and resulted in a more conservative estimate of the number of limited English proficient children (see Miranda, 1980; p. 6 for further information). However, the data in table A-4 did not reflect this conservative approach. To remedy this situation, Ms. Miranda has submitted a revised table A-4 for inclusion in the final NIE report on CESS. A copy of the table is in appendix C.

In developing the LM&AI, five techniques were proposed as alternatives for determining the critical scores. The five techniques[33] are summarized below:

(1) For each age-level test determine the score which (on Field Test II data) was one standard deviation below the mean score for the FES (Fluent English Speakers) group of that age

(2) Similarly, use that score which was one standard deviation above the mean score for the LESA (later revised to LEP) group of each age

(3) Use the highest Field Test III LESA (LEP) score made by any individual on each age group test

(4) Plot the score of LESA and FES separately and select the score equivalent to the point of intersection of the two distributions

(5) Use discriminant function analysis (DFA) which considers subscores to determine a centroid, which can act as the critical point.

After examining the "accuracy" of the various alternatives, DFA was chosen as the method for determining the critical scores.

While NCES/ORA fully endorses the use of DFA, there remains an issue regarding its use which concerns us. This concerns the application of DFA without concern for the differential "costs" of misclassification. DFA is a very powerful tool in that it minimizes the total proportion of the sample which is misclassified. However, if the resulting classification criteria (critical scores) consistently misclassify one subgroup (e.g., LEP) at the expense of the other, a serious bias may result. More explicitly, if there are actually N_1 LEP children and N_2 English proficient (fluent) children among the $N = N_1 + N_2$ children of non-English language background households, then the cutoff score will lead to an unbiased classification procedure if and only if $N_2 \cdot Pr(\text{Classified LEP} \mid \text{Actually fluent}) = N_1 \cdot Pr(\text{Classified fluent} \mid \text{Actually LEP})$. That is, the expected number of fluent children misclassified as LEP must equal the expected number of LEP children misclassified as fluent.

In defense of the procedures used, since N_1 and N_2 were not known in advance, minimizing the overall misclassification error makes reasonable sense. However, as can be seen in table A, the actual discrimination procedure used was much more likely to misclassify LEP children than fluent children. This explains why the critical scores for DFA seemed low.[34]

Table A presents the estimated conditional probabilities of correct and incorrect classifications by the LM&AI for the critical scores found in the revised table A-4 (appendix C).

Table A: Estimated conditional probabilities of correct and Incorrect Classifications by the LM&AI (See appendix C revised table A-4)

Age	P_{11} [1]	P_{12} [2]	P_{21} [3]	P_{22} [4]
5	0.811	0.000	0.189	1.000
6	0.795	0.000	0.205	1.000
7	0.806	0.000	0.194	1.000
8	0.893	0.000	0.107	1.000
9	0.813	0.000	0.188	1.000
10	0.833	0.000	0.167	1.000
11	0.682	0.000	0.318	1.000
12	0.864	0.182	0.136	0.818
13	0.800	0.000	0.200	1.000
14	0.879	0.204	0.121	0.796

[1] P_{11} = Pr(Classified LEP|Actually LEP).

[2] P_{12} = Pr(Classified LEP|Actually Fluent).

[3] P_{21} = Pr(Classified Fluent|Actually LEP).

[4] P_{22} = Pr(Classified Fluent|Actually Fluent).

The bias evident in table A led NCES/ORA to conclude that the critical scores for each age level test of the LM&AI should be revised in order to remove the estimated bias, once we have computed estimates of N_1 and N_2. The mechanism by which this can be done follows:[35]

> Let P_{11}, P_{12}, P_{21}, and P_{22} be defined as they are found in table A. Let N_1 and N_2 be the actual number of LEP and fluent children, respectively. Finally, let L and F be the expected number of LEP and fluent children, respectively, as estimated by the LM&AI. Then,

$$L = N_1 P_{11} + N_2 P_{12}$$

$$F = N_1 P_{21} + N_2 P_{22}$$

Solving for N_1 and N_2, we get

$$N_1 = (LP_{22} - FP_{12})/(P_{11} P_{22} - P_{12} P_{21}) \text{ and}$$

$$N_2 = (FP_{11} - LP_{21})/(P_{11} P_{22} - P_{12} P_{21})$$

Of course, the values of P_{11}, P_{12}, P_{21}, and P_{22} are functions of the actual critical scores which are used for separating LEP from fluent children. This means that an iterative procedure must be used to determine the unbiased estimates of N_1 and N_2 based on critical scores associated with "balanced" misclassification errors. To accomplish this, the estimated "misclassification balance", defined by $|L \cdot P_{21} - F \cdot P_{12}|$, must be calculated for each possible critical score. For each age group, the critical score is selected which minimizes the estimated misclassification imbalance. Using the expected number of LEP (L) and fluent (F) children and the revised probabilities (P's) once the expected misclassification imbalance has been minimized, we can approximate the "unbiased" values of N_1 and N_2 for each age group. The values of L, F, P_{11}, P_{12}, P_{21} and P_{22} which were used to compute N_1 and N_2 are found in appendix D.

The results of these computations (shown in table B) clearly demonstrate the consistent bias in the LM&AI classification procedure. The CESS/LM&AI LEP counts underestimate the "true" values at every age, except for ages 12 and 14.

Table B: Effect on 1978 CESS LEP counts of removing the estimated bias

Age	1978 CESS LEP Count	"Unbiased" LEP Count
Total	2,408,875	2,621,332
5	192,297	249,734
6	291,622	306,970
7	275,924	320,774
8	257,807	277,422
9	167,304	189,277
10	294,156	329,047
11	190,064	266,706
12	251,680	207,388
13	196,577	227,732
14	291,444	246,282

In actual practice, the LEP counts determined by the critical score will almost always differ from the "unbiased" estimate, since all children with a given score must fall on one side of the critical score or the other. Therefore, we must accept some bias in our counts, but NCES/ORA has minimized the expected bias by using the procedure just described. Table C contains the CESS Draft Report critical score (with the slight modification mentioned earlier), the revised critical scores, and the resulting LEP count for each age level. Note that the national LEP figure of 2,631,075 (table C) compares to an "unbiased" estimate of 2,621,332 (table B).

Table C: Revised critical scores and resulting LEP counts

AGE	CESS Draft Report critical score	Revised critical score	Revised LEP count
Total	___	___	2,631,075
5	18.5	25.5	254,657
6	26.5	29.5	303,584
7	39.5	44.5	318.470
8	38.5	40.5	280,256
9	43.5	46.5	188,187
10	49.5	52.5	330,979
11	41.5	51.5	271,485
12	46.5	44.5	208,426
13	48.5	52.5	229,986
14	52.5	49.5	245,045

By minimizing the estimated bias, a less conservative, yet more analytically sound, LEP count results with a change in the National CESS estimate from 2,403,375 to 2,631,075 LEP children. This change represents a National increase of 9.22 percent in the number of LEP children estimated in the CESS Draft Report.

Recommendation NCES/ORA recommends that the NIE final report on CESS reflect this analysis and the revised LEP counts found in table C.

ISSUE: What were the effects of non-response bias on the counts and estimate of the number of LEP children?

Discussion The question to be addressed is whether nonrespondents are similar to or different from respondents to the study. There is no evidence in the NIE Draft Report of November 1979 to indicate that nonresponse bias was empirically investigated.

Dr. Donald Rogers, Vice President for Operations with Resource Development Institute (one of the subcontractors) completed and forwarded to the author a brief paper in response to AI/CESS. In his paper, Dr. Rogers presents: "The results of a very, very simple analysis of the effects of nonresponse during the CESS study."[36] A copy of Dr. Rogers' paper is in appendix B.

Dr. Rogers' stated in a letter that accompanied his paper that:

> My assumptions [appendix B] generated a weighted LESA [LEP] total that fell within the 95 percent confidence interval for the total weighted U.S. LESA [LEP] count reported by the CESS study. I do not believe that a study of nonrespondents will greatly increase or decrease the total, weighted U.S. LESA [LEP] count.[37]

Recommendation NCES/ORA concurs with Dr. Rogers' position that further investigations of nonresponse bias associated with the 1978 CESS are not warranted.

NOTES

1. Dubois, 1980; p. 1.

2. Ibid., pp. 2-3.

3. Sterling and Weinkam, 1979; p. 2.

4. Ibid., p. 12.

5. Troike (1980) suggests that a restatement of this issue should not detract attention from the fact that there is a dearth of and, therefore, a need for basic research on the question of which types of items are appropriate for language assessment and measurement at each age level. For example, it is necessary to examine the range of grammatical or semantic variations which are tolerable for each test item at each age level. Only after examining this question and others, says Troike, can we hope to be confident of obtaining reliable and valid measures of language proficiency.

6. Miranda, 1980; p. 2.

7. Ibid.

8. O'Malley, 1980; p. 2.

9. Ibid.

10. Ibid., p. 1.

11. Ibid., p. 2.

12. Ibid.

13. Dubois, 1980; p. 6.

14. O'Malley, 1980; p. 3.

15. DeAvila and Duncan, 1976; p. 9.

16. Ibid.

17. Ibid.

18. Dubois, 1980; p. 7.

19. Cummins, 1979; p. 227.

20. Dubois, 1980; p. 6.

21. Miranda, 1980; p. 2.

22. DeAvila, Duncan, Ulibarri, and Fleming, June 1979; p. 8.

23. Ibid., p. 50.

24. Cazden, 1972; p. 226.

25. Ibid.

26. Cazden, 1972; pp. 230-232. DeAvila, Duncan, Ulibarri, and Fleming, June 1979; p. 51.

27. Cummins, 1979; p. 229.

28. DeAvila, Duncan, Ulibarri, and Fleming, June 1979; p. 53.

29. Ibid.

30. Ibid.

31. Ibid., p. 38.

32. Tucker, 1979; p. 75.

33. Miranda, 1979; p. 38.

34. Ibid., p. 43.

35. The analyses presented here were developed by Dr. Rolf M. Wulfsberg, Assistant Administrator for Research and Analysis, NCES.

36. Rogers, 1980; appendix B, p. 1.

37. Ibid., p. 1.

Bibliography

Cazden, Courtney B. Child Language and Education. New York: Holt, Rinehart and Winston, Inc., 1972.

Cummins, James. "Linguistic Interdependence and the Educational Development of Bilingual Children." Review of Educational Research, vol. 49, no. 2, Spring 1979, pages 222-251.

DeAvila, Edward A. and Sharon E. Duncan. "A Few Thoughts About Language Assessment: The Lau Decision Reconsidered." Larkspur, California: DeAvila, Duncan & Assoc., June 17, 1976, 33 pages (photocopy).

DeAvila, Edward and Sharon Duncan. Relative Linguistic Proficiency and Field Dependence/Independence: Some Findings on Linguistic Heterogeniety and Cognitive Style of Bilingual Children. Austin, Texas: Southwest Educational Development Laboratory, February 1979.

DeAvila, Edward A., Sharon E. Duncan, Daniel M. Ulibarri, James S. Fleming. Predicting the Success of Language Minority Students from Developmental, Cognitive Style, Linguistic, and Teacher Perception Measures. Austin, Texas: Southwest Educational Development Laboratory, June 1979.

Dubois, David D. "Analytical Issues Regarding the Children's English and Services Study." Washington, D.C. : U.S. Department of Education, National Center for Education Statistics, January 21, 1980, 16 pages (photocopy).

Edmonds, Marilyn H. "New Directions in Theories of Language Acquisition." Harvard Educational Review, vol. 46, no. 2, May 1976, page 175.

Lambert, W. E. "Culture and Language as Factors in Learning and Education." Education of Immigrant Students (A. Wolfgang, Ed.) Toronto: Ontario Institute for Studies in Education, 1975.

Martinez, Jose. "Response to Issues on LM&AI." Response to AI/CESS with cover letter dated February 22, 1980 at Sacramento, California, 5 pages.

Miranda, Lourdes. CHILDREN'S ENGLISH AND SERVICES STUDY: Technical Report on the LM&AI. Bethesda, Maryland: L. Miranda and Associates, Inc., September 10, 1979, 44 pages.

Miranda, Lourdes and Associates. "Response to the Office of Research and Analysis of the National Center for Education Statistics' Inquiry on Three Analytical Issues Associated with the 1978 Children's English and Services Study." Bethesda, Maryland: L. Miranda and Associates, Inc., February 15, 1980, 10 pages.

O'Malley, J. Michael. Language Minority Children with Limited English Proficiency in the United States, Spring, 1978 (Draft Report). Washington, D.C.: National Institute of Education, 1979.

O'Malley, J. Michael. Response to AI/CESS in the form of a letter dated January 31, 1980 at Alexandria, Virginia, 4 pages.

Peng, Samuel S. Response to AI/CESS in the form of a letter dated February 12, 1980 at Rockville, Maryland, 5 pages.

Rogers, Donald D. Response to AI/CESS in the form of a letter and attachment dated February 3, 1980 at Austin, Texas, 5 pages.

Rowlett, Karen. Response to AI/CESS in the form of a letter dated February 8, 1980 at Austin, Texas, 3 pages.

Silverman, Leslie J. Response to AI/CESS in the form of an (HEW) Education Division memorandum dated February 7, 1980 at Washington, D.C.

Sterling, T. D., J. J. Weinkam. "What Happens When Major Errors are Discovered Long After An Important Report Has Been Published?" Paper presented to the American Statistical Association Annual Meeting in Washington, D.C. on August 16, 1979. Burnaby, British Columbia: Simon Fraser University, 1979, 13 pages (photocopy).

Tremaine, R. V. Syntax and Piagetian Operational Thought. Washington, D.C.: Georgetown University Press, 1975.

Troike, Rudolph C. Personal communications with Dr. David D. Dubois, May 30, 1980.

Truex, Kathy. Response to AI/CESS in the form of an (HEW) Office of the Secretary memorandum dated March 18, 1980 at Washington, D.C.

Tucker, G. Richard. "Bilingual Education: Some Perplexing Observations." Educational Evaluation and Policy Analysis, vol. 1, no. 5, September-October, 1979, pp. 74-75.

Whorf, B. L. Language, thought and reality: Selected Writings of Benjamin Lee Whorf. (J. B. Carroll, Ed.) Cambridge: M.I.T. Press, 1956.

APPENDIXES

A: Analytical Issues Regarding the Children's English and Services Study

ANALYTICAL ISSUES REGARDING THE CHILDREN'S ENGLISH AND SERVICES STUDY

Prepared by
Dr. David D. Dubois, Policy Analyst
Office of Research and Analysis
National Center for Education Statistics
(January 21, 1980)

Introduction

The 1978 Child[illegible]'s English and Services Study (CESS) was recently completed under contract from the National Institute of Education (NIE), with shared support from the National Center for Education Statistics (NCES) and the U.S. Office of Education (USOE). The final project report will be published by NIE. The principal objective of the CESS was to objectively determine an estimate of the number of limited English proficient (LEP) children between the ages of 5 and 14 in the United States.

The purpose of this paper is to present and discuss three analytical issues which have been identified as a result of an assessment of the research design, data analyses and other information which are described in the 1978 CESS draft report of September 6, 1979 (and a later revision dated November, 1979) entitled "Language Minority Children With Limited English Proficiency in the United States: Spring 1978."

This inquiry is sponsored by the Office of Research and Analysis (ORA) of the National Center for Education Statistics (NCES). To date, reviewers have included the NCES Assistant Administrator for Research and Analysis, the ORA Policy Analyst, and an external consultant from the American Institute for Research in the Behavioral Sciences (AIR) whose services were obtained under contract with the NCES/AIR Statistical Analysis Group in Education. This paper is based entirely upon these reviews.

Recipients of this paper are invited to respond to the analytical issues. Based upon their responses and other information, the Office of Research and Analysis will publish a position paper on the resolution of the identified analytical issues.

Objective of the Inquiry

From an analytical point of view, the 1978 CESS could become a landmark in the determination of estimates of the number of LEP children in the United States. The CESS estimate of the number of LEP children was accomplished directly by developing and administering a domain-referenced content test to a sample of children from language minority households in order to assess English language skills in speaking, reading, writing, and understanding. Prior to 1978, estimates of this type were derived by using surrogate or indirect measures.

It is anticipated that the results of the 1978 CESS will be extensively used and frequently cited by U.S. Government officials, members of the U.S. Congress, and others. At NCES, for example, it is anticipated that the CESS data base will be used, with other surrogate measures, to calibrate the 1980 U.S. Census data in order to determine recent and accurate LEP person counts. Additionally, the CESS data base will be a component data base of the NCES study to determine projections of the numbers of LEP persons in the U.S. for the next 5, 10, 15, and 20 years.

Since the results of the 1978 CESS are of tremendous importance to present and future research studies, bilingual program and policy development, and funding for bilingual education, unresolved analytical issues which could adversely affect the validity of the results are being stated with the hope

of their resolution. As mentioned earlier, the ORA will publish a position paper as a result of this inquiry. In the position paper, we expect to provide a technical reply to each issue. Our reply is expected to include recommendations or suggestions for additional research tasks and/or caveats to current CESS reports which could, in our opinion, improve the quality of the products we now have.

Invitation to Respond

Recipients of this paper are encouraged to respond to the issues. Respondents are assured that their contributions will be carefully considered prior to the development and issuance of the ORA position paper. The position paper will be released only after each recipient (or his or her designate) has responded or has indicated that he or she will not respond to the issues.

Written replies must be received no later than the close of business, Friday, February 8, 1980. Replies to the issues must be written and should be addressed to:

Dr. David D. Dubois, Policy Analyst
National Center for Education Statistics
400 Maryland Avenue, SW, Room 3153
Washington, DC 20202
(Telephone: 202-245-8233)

The persons listed below were designated to receive a copy of this paper.

Name	Agency
Edward Bryant	Westat, Inc.
Lois-Ellin Datta	National Institute of Education
Karen Dietz/Don Rogers	Univ. of Texas-Austin/Resource Development Institute
Josué M. Gonzales	Office of Bilingual Education, OE

Ron Hall	Office of the Assistant Secretary for Education (Policy Development)
Ty Hartwell	Research Triangle Institute
Reynaldo Macias	National Institute of Education
Jose Martinez	California State Department of Education
Lourdes Miranda-King	L. Miranda and Associates, Inc.
J. Michael O'Malley	National Institute of Education
Samuel Peng	Westat, I-
Leslie Silverman	National Center for Education Statistics
Kathy Truex	Office of the Assistant Secretary for Planning and Evaluation
James Vanecko	Office of the Assistant Secretary for Education (Policy Development)
Carl Wisler	Office of Evaluation and Dissemination, OE

History of the 1978 CESS

The 1978 CESS was developed by NIE through a contract with L. Miranda and Associates, Inc. Lourdes Miranda-King was the project director. Dr. J. Michael O'Malley was the NIE project officer and Leslie J. Silverman was the NCES coordinator. Subcontractors included Westat, Resource Development Institute and Research Triangle Institute.

The primary mission of the 1978 CESS was to objectively determine an estimate of the number of LEP children, ages 5-14, inclusive, in the United States. A nationally representative sample of households was surveyed during the Spring of 1978. Households were identified where a language other than English was spoken and where children between the ages of 5 and 14 were living. The Language Measurement and Assessment Inventory (LM&AI), a test in English that determines whether or not a child is limited in English language proficiency,

was developed. Selected children from the identified households were individually administered the LM&AI. Specifications for the survey design and the LM&AI were provided by an advisory group composed of State Education Agency representatives in bilingual education, assessment, and data collection.

The LM&AI was designed to measure skills in speaking, understanding, reading and writing in English. The test is domain-referenced for objectives that children at ages 5-14 would be expected to perform in order to profit from instruction in an all-English language educational environment.

Ten separate tests, one for each age, were developed and used in the survey. Reliabilities of the test for the separate forms range from .86 to .92. As a result of preliminary field tests of the LM&AI, a critical score for each age test was determined which could be used to classify each child as proficient in English or as limited English proficient.

The contractor provided three cautionary caveats regarding the LM&AI. First, the LM&AI was not designed to determine placement or diagnosis with individual children in educational settings. Second, the instrument was designed in a manner that resulted in an unknown level of cultural bias. Third, the LM&AI items are not "pure" measures of English language proficiency; some of the items assess English language proficiency, memory and cognitive ability.

We understand that the final NIE report of the 1978 CESS is scheduled for publication in the immediate future.

Statement and Discussion of the Issues

Three analytical issues are presented and discussed. They are:

1. Were the items which were selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?

2. Were the cutoff scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency, set properly?

3. What were the effects of non-response bias on the counts and estimates of the number of LEP children?

If the first question is answered negatively, then the value of the entire 1978 CESS is brought into question. In the event that it is answered affirmatively, then a negative answer to the second question would imply the need for further analyses of the CESS data -- and possibly the collection of additional data -- in order to re-compute the cutoff scores. The issue raised by the third question could be empirically investigated in the event that it was decided to collect the additional data described earlier.

A detailed discussion of each issue follows.

ISSUE: Were the items which were selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?

Discussion. Each age-level instrument of the LM&AI consisted of a set of items that could be scored so that a high score would indicate that the child was proficient in English while a low score would indicate that the child was limited English proficient. Therefore, the issue can be rephrased in the following manner: Is English language proficiency the dimension on which the scores vary, or are other dimensions associated with variation in the scores? More specifically:

- Are the test scores related to language dominance?
- Are the test scores related to general language development?

The question of language dominance is addressed in the project Draft Report (November, 1979):

> English should be the exclusive criterion irrespective of the child's proficiency in the non-English language. Thus, language dominance was considered irrelevant to the discussion. (Page II-3)

This objective of the study is subject to question on the basis that for bilingual education policy development, a child's dominant language might affect the potential benefits from participation in a bilingual education program. The reader is cautioned that this review does not attempt to operationally define the phrase "bilingual education program" and that this omission was intentional. Whether it is appropriate to assess English language proficiency, ignoring the child's proficiency in another language, remains a policy question to the addressed.

Are the scores on the test related to general language development? The project Draft Report (November, 1979) states that:

> . . . items on the test are not "pure" measures of English language proficiency. In some cases, the items assess English language proficiency, memory, and cognitive ability. The intermingling of the potentially disparate constructs was intentional to give the items as much validity for representing important school tasks as possible. (Page A-10)

Any test so developed could also differentiate between two children with *equal* English language proficiency, giving a higher score to the child with greater memory and/or cognitive abilities. It could be argued, therefore, that the test development procedures should have excluded items not primarily associated with English language proficiency. The types of items selected for the test (Draft Report; November, 1979; Table A-1) appear to be generally assessing relevant content. There is, however, a component of general cognitive development, not merely English language development.

The choice of items for the LM&AI was a function of a field test. Items were selected that best differentiated between two criterion groups. The project Draft Report (November, 1979) states:

> The test was being developed to differentiate language minorities who were limited in English proficiency from those who could profit from instruction in English. Items under development were to be field tested with two clearly defined criterion groups: (a) limited English proficient children; and (b) fluent English speaking children who were clearly profiting from instruction in English. (Page II-6)

The test was clearly being prepared for administration to language minority children. The dimension being tested is essentially the dimension on which those two groups differ most. It could be argued that the two groups differed on native language as well as English language proficiency and, therefore, the test scores could be expected to have a partial language dominance loading. A potential solution to this problem would be to equate the two criterion groups on proficiency in a non-English language. This would make the test independent of language dominance.

ISSUE: Were the cutoff scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency, set properly?

Discussion. The purpose of the LM&AI was to provide a mechanism for making a dichotomous assignment of a child as being either English proficient or limited English proficient. Therefore, the cutoff score which was chosen for each age-level test of the LM&AI is critical for the determination of valid counts. As an example, if the cutoff score on each age-level test is lowered by two items, the estimated count of LEP children decreases from 2.41 million to 2.13 million children, or a decrease of 280,000. Similarly, if the cutoff score for each age-level test is raised by two items, the estimated count of LEP children is increased from 2.41 million to 2.62 million children, or an increase

of 210,000. Thus, a score difference of four items has the effect of altering the count by nearly one-half million.

Recall that the cutoff score was that score which best differentiated LEP children from fluent English-speaking (FES) children who were clearly profiting from instruction in English.

In developing the LM&AI, five techniques were proposed as alternatives for determining the cutoff scores. The five techniques are summarized on page 38 of the CHILDREN'S ENGLISH AND SERVICES STUDY: Technical Report on the LM&AI (L. Miranda and Associates, Inc., September 10, 1979):

(1) For each age-level test determine the score which (on Field Test II data) was one standard deviation below the mean score for the FES group of that age.

(2) Similarly, use that score which was one standard deviation above the mean score for the LESA (later revised to LEP) group of each age.

(3) Use the highest Field Test III LESA score made by any individual on each age group test.

(4) Plot the scores of LESA and FES separately and select the score equivalent to the point of intersection of the two distributions.

(5) Use discriminant function analysis (DFA) which considers subscores to determine a centroid point, which can act as the critical point.

After examining the "accuracy" of the various alternatives, DFA was chosen as the method for determining the cutoff scores. While this Office fully endorses this choice, there remain three subissues which still bother us.[1]

First, the above excerpt from the Technical Report implies that subscores were used in the DFA. If this is so, several events must have happened.

---

[1]The analyses found here were developed by Dr. Rolf M. Wulfsberg, the Assistant Administrator for Research and Analysis at NCES.

(1) The subscores would be transformed into a new total score representing a linear combination of the subscores. This new score would be real-valued (as opposed to integer-valued) and it would be conceivable -- in fact, highly likely -- that relative scores between two individuals could be reversed. That is, if individual A had a higher original score than individual B, the revised DFA score for A could easily be lower than that of B due to differential weighting of the subscores. Since no scores on the final CESS tape are non-integer-valued, and since no reversal of the kind discussed above occurred, one can only assume that subscores were, in fact, not used in the DFA.

(2) The relative weighting of the items, which was carefully designed, would be totally revised by the differential weighting of the DFA procedure. This is another reason that this Office doubts that subscores were used.

The second subissue concerns the application of DFA without concern for the differential "costs" of misclassification. DFA is a very powerful tool in that it minimizes the total proportion of the sample which is misclassified. However, if the resulting classification criteria (cut scores) consistently misclassify one subgroup (e.g., LEP) at the expense of the other, a serious bias may result. More explicitly, if there are actually N_1 LEP children and N_2 English proficient (fluent) children among the $N = N_1 + N_2$ children of non-English language background households, then the cutoff score will lead to an unbiased classification procedure if and only if N_2.Pr(Classified LEP/Actually fluent) = N_1.Pr(Classified fluent/Actually LEP). That is, the expected number of fluent children misclassified as LEP must equal the expected number of LEP children misclassified as fluent.

In defense of the procedures used, since N_1 and N_2 were not known *a priori*, minimizing the overall misclassification error makes reasonable sense. However, as can be seen in Table A, the actual discrimination procedure used was much more likely to misclassify LEP children than fluent children. This explains why the cut scores for DFA seemed low (see page 43 of the aforementioned Technical Report).

Table A: Estimated Conditional Probabilities of Correct and Incorrect Classifications by the LM&AI

Age	p_{11} [1]	p_{12} [2]	p_{21} [3]	p_{22} [4]
5	0.892	0.000	0.108	1.000
6	0.955	0.037	0.045	0.963
7	0.889	0.000	0.111	1.000
8	0.929	0.000	0.071	1.000
9	0.906	0.000	0.094	1.000
10	0.944	0.000	0.056	1.000
11	0.795	0.000	0.205	1.000
12	0.864	0.182	0.136	0.818
13	0.880	0.000	0.120	1.000
14	0.879	0.204	0.121	0.796

[1] P_{11} = Pr(Classified LEP|Actually LEP).

[2] P_{12} = Pr(Classified LEP|Actually Fluent).

[3] P_{21} = Pr(Classified Fluent|Actually LEP).

[4] P_{22} = Pr(Classified Fluent|Actually Fluent).

The evident bias described above raises the third subissue: Should the cutoff scores be revised to remove the estimated bias after the fact (when we have estimates of N_1 and N_2)? This Office tends to feel that this should be done. The mechanism by which this could be done is described below.

Let P_{11}, P_{12}, P_{21}, and P_{22} be defined as in Table A. Let N_1 and N_2 be the actual number of LEP and fluent children, respectively. Finally, let L and F be the expected number of LEP and fluent children, respectively, estimated by the LM&AI. Then,

$$L = N_1P_{11} + N_2P_{12} \quad (1)$$
$$F = N_1P_{21} + N_2P_{22} \quad (2).$$

Solving for N_1 and N_2, we get

$$N_1 = (LP_{22} - FP_{12})/(P_{11}P_{22} - P_{12}P_{21}) \text{ and}$$
$$N_2 = (FP_{11} - LP_{21})/(P_{11}P_{22} - P_{12}P_{21}).$$

By using the actual CESS estimates for L and F, we can then approximate the unbiased values of N_1 and N_2 for each age group. The results, which are shown in Table B, clearly demonstrate the consistent bias in the LM&AI classification procedure. *The CESS/LM&AI LEP counts underestimate the "true" values at every age, except for age 14.*

Table B: Effect on LEP Counts of Removing Estimated Bias

AGE	CESS LEP COUNT	"UNBIASED" LEP COUNT
5	192,297	215,580
6	291,622	301,767
7	275,924	310,375
8	257,807	277,510
9	167,304	184,662
10	294,156	311,606
11	190,064	239,074
12	251,680	262,412
13	196,577	223,383
14	291,444	284,766
Total	2,408,875	2,611,135

If we accept the new LEP counts as more realistic estimates of the true values, then we can adjust the cut scores to reflect these new counts by raising (except for age 14) the cut scores until the proper number of children have been classified as LEP. In reality, this point will (almost) always fall in the middle of a cell (score), so one can choose the cut score which will yield the closest estimate to N_1.

In the case which is present in Table C, a different rule was used. Since there is an abnormal "roller coaster" effect to the data to begin with in the relationship between age and percent LEP, the cut score leading to the percentage closest to the overall mean percentage was chosen for each age group. That is, the lower cut score was generally used for even ages and the higher cut score was generally used for odd ages.

Table C: Modified Cut Scores and Resulting LEP Counts

Age	Old cut score	New cut score	New LEP count
5	18.5	21.5	223,327
6	26.5	28.5	298,929
7	39.5	43.5	307,759
8	38.5	39.5	268,830
9	43.5	46.5	188,187
10	49.5	50.5	310,860
11	41.5	45.5	246,921
12	46.5	46.5	251.680
13	48.5	51.5	223,785
14	52.5	52.5	291,444
Total	-	-	2,611,722

The relationships among the CESS/LM&AI estimates, the unbiased estimates, and the adjusted CESS estimates are evident in Chart 1. Chart 1 shows the percent of each age cohort for each of the estimates of the number of limited English proficient children.

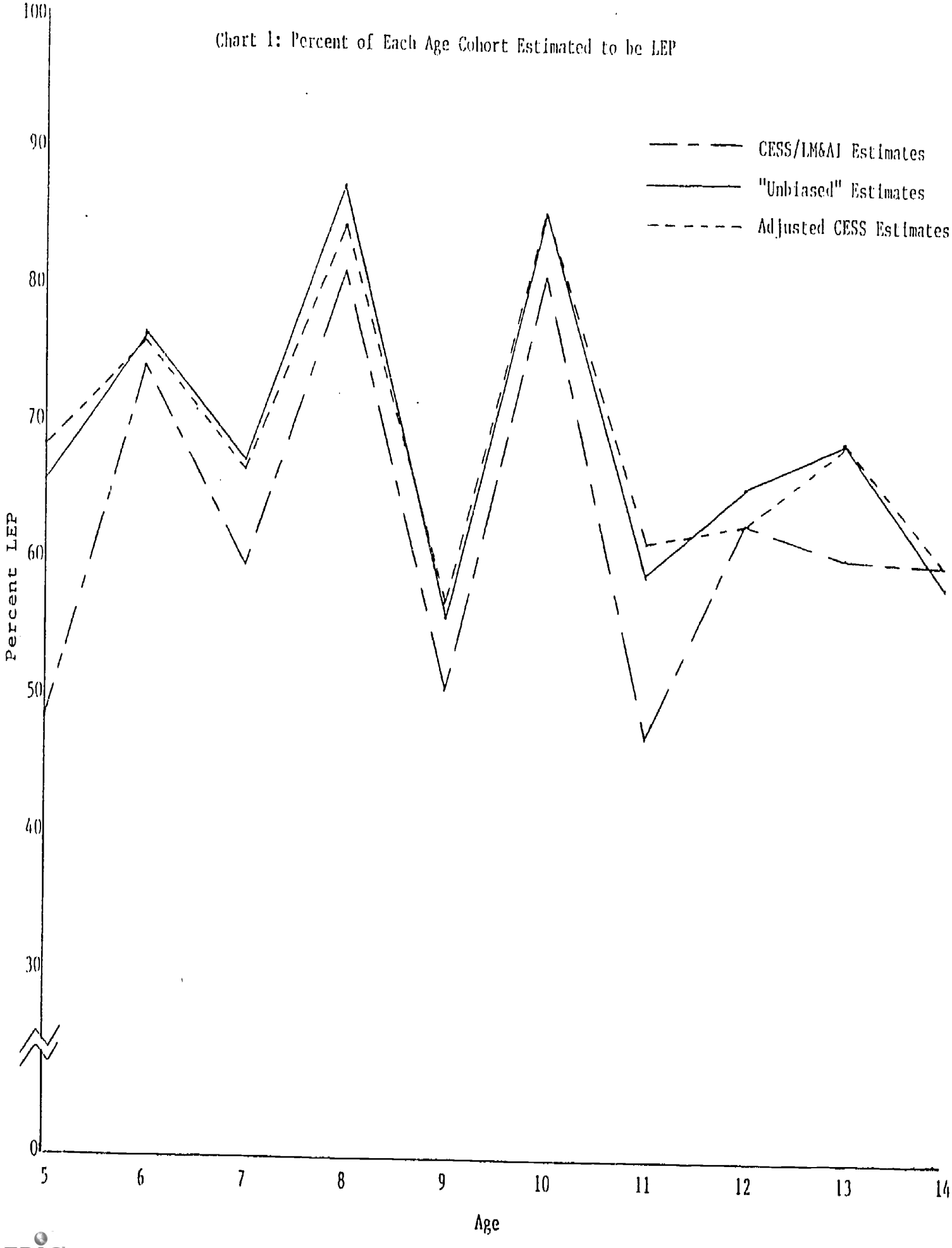


Chart 1: Percent of Each Age Cohort Estimated to be LEP
CESS/LM&AI Estimates
"Unbiased" Estimates
Adjusted CESS Estimates
Percent LEP
100
90
80
70
60
50
40
30
0
5
6
7
8
9
10
11
12
13
14
Age

The procedure described above should go a long way toward removing the bias in the LM&AI. Of course, the values of P_{ij} used in the derivation are conditioned on the original cut scores used by the LM&AI. With the modified cut scores, the P_{ij}'s would change (as do the new L and F counts shown in Table C), so that the results could still change slightly. (This is because the LM&AI sample of 354 fluent children and 337 LEP children are not necessarily representative of their respective populations.) The Office of Research and Analysis is attempting to obtain the original data which were used to determine the cutoff scores on the LM&AI from L. Miranda and Associates, Inc. in order to explore this issue.

ISSUE: What were the effects of non-response bias on the counts and estimate of the number of LEP children?

Discussion. In survey research of this type, the potential effects of non-response bias are a reality. The question to be addressed is whether non-respondents are similar to or different from respondents to the study.

Response rates by regional subpopulations (New York, Texas, California, remainder of the U.S.) for the household screener, questionnaire and for the administration of the LM&AI are presented in Table III-1 of the Draft Report (November, 1979). From the table it can be determined that the response rates, totaled over all subpopulations, were: household screener, 76.2%; household questionnaire, 93.8%; LM&AI administration, 84.6%. Response rates were derived by using the formula

$$\text{Response Rate} = \frac{\text{Total Number Completed}}{\text{Total Number Eligible}} \times 100.$$

There is no evidence in the Draft Report (November, 1979) to indicate that non-response bias was empirically investigated. Although adjusting weights by poststratification is a customary practice, it can be argued that this is not

necessarily a satisfactory substitute for empirically investigating differences between respondents and non-respondents.

In the event that the first issue stated herein is answered in the affirmative and, additionally, a decision is made to collect additional data for recalibration of the LM&AI, an empirical investigation of non-response bias can be undertaken concurrently.

In summary, ORA reviewers believe that these issues can be resolved and, accordingly, that the study can be retained by cooperative responsible action.

B: Rogers' Paper, Nonresponse Analysis

AUTHOR: Dr. Donald Rogers
Resource Development Institute
Austin, Texas

ATTACHMENT A: Nonresponse Analysis

Purpose

The purpose of this paper is to present the results of a very, very simple analysis of the effects of nonresponse during the CESS study.

General Procedure

The general procedure was to assume that nonresponding "SCR incomplete; probable ineligible households (Code 8 households)" had characteristics that were significantly different from responding households. The impact of this assumption was then determined by reweighting the data and recomputing NELB and LESA counts.

Limitations

The analysis reported here uses average weights. Ideally, each stratum is considered individually. However, the resources required for a stratum-by-stratum analysis were not available. Therefore average weights were used because they were easy to compute. This means that the results of this analysis only indicate or suggest the type of results that would be obtained by a sophisticated analysis.

References

This paper is based on the information presented in RDI's final CESS reports. Data have been taken from Section 8 (Data Analysis Procedures) and Section 9 (Results) of Volume I. Weighting formulae are taken from Appendix 6.6 of Volume II. The reader must have these reports to follow this paper. For example, the definitions of variables are presented in Appendix 6.6 and are not repeated here.

Assumptions

The following assumptions were made to assess the effects of nonresponse:

1. All Category 8 households complete the SCR.

2. The percentage of Category 8 households that are eligible and complete the HHQ is twice as great as the percentage of Category 1, 2, and 3 households.

3. All of the eligible Category 8 households complete the HHQ.

4. The average number sampled per eligible household is the same for Category 8 households.

5. The average number of completed LM&AI per household is the same for Category 8 households.

6. The average number of LESA children per household is the same for Category 8 households.

The effects of these assumptions on the "raw" data is presented in the following tables.

Household Codes	Number Complete SCR	Percent Complete HHQ	Number Complete HHQ
1,2,&3	25,358	6.5	1,652
8	5,790	13.0	753
Totals	31,148		2,405

Household Codes	Number Complete HHQ	Average Number Sampled Per Household	Number Sampled	Average Number of Completed LM&AI Per Household	Number of Completed LM&AI
1,2,&3	1,652	1.78	2,953	1.16	1,909
8	753	1.78	1,340	1.16	873
Totals	2,405		4,293		2,782

Household Codes	Number of Completed LM&AI	Percent LESA	Number LESA
1,2,&3	1,909	71.24	1,360
8	873	71.24	622
Totals	2,782		1,982

Changes in Values

The raw data were used to compute average values for the variables in Appendix 6.6. The computed values are presented in the table below. The formulae have been omitted because they appear in Appendix 6.6. Although the use of the symbols is not entirely appropriate and is not precisely consistent with the definitions presented in Appendix 6.6, the results are presented in this manner to make it easy for the reader to follow the calculations.

Variable	Average or Estimated Value Before Assumptions	Average or Estimated Value After Assumptions
r_{hij}	33,283	33,283
r'_{hij}	25,358	31,148
W'_{hij}	64.5	64.5
$\hat{L}_{hs}$	2,146,753	2,146,753
$\hat{L}'_{hs}$	1,635,591	2,009,046
$W^{(1)}_{hij}$	83.9	68.9
S_{hij}	1,762	2,515
S'_{hij}	1,652	2,405
$\hat{M}_{hs}$	147,832	173,284
$\hat{M}_{hs}$	138,603	165,704
$W^{(2)}_{hij}$	89.5	72.1
$\Sigma Q'_{hijm}$	3,084,452	3,048,452
C'_{hijm}	2,953	4,293

Variable	Average or Estimated Value Before Assumptions	Average or Estimated Value After Assumptions
C_{hijm}	34,061	49,573
Q'_{hijm}	1,032	833
C''_{hijm}	1,909	2,782
$\hat{N}_{hs}$	3,047,496	3,571,776
$\hat{N}'_{hs}$	1,970,088	2,314,624
Q_{hijm}	1,597	1,283
Q(adjusted for SIE)	1,997	1,370
Total NELB	3,811,850	3,811,850

Analysis

The assumptions about the Category 8 households increased the sampled number of NELBs from 1,909 to 2,782. This is approximately a 46% increase. However, because of the weighting procedures, this increase has no meaningful effect on the total U.S. estimates.

The assumption about the Category 8 households increased the sampled number of LESAs from 1,360 to 1,982. This is approximately a 46% increase. The effects on the total U.S. estimate depends upon assumptions about how these cases are weighted. The table presented below reports the average weights that have been used to this point in the analysis.

Assumption	Type	Number	Average Weight	U.S. Estimate
Before	NELB	1,909	1,997	3,811,850
After	NELB	2,782	1,370	3,811,850
Before	LESA	1,360	1,771	2,408,908
After	LESA	1,982	Unknown	Unknown

This table indicates that before the assumptions, the average LESA weight is less than the average NELB weight. The assumptions that have been made should not affect this relationship, and the average LESA weight should continue to be less than the average NELB weight. However, to test response bias, assume the NELB and LESA average weights are the same after the assumptions and are equal to 1,370. This yields a total U.S. estimate of 2,715,340 LESAs. This estimate is 306,432 LESAs greater than the LESA estimate reported by the CESS study. However, an estimate of 2,715,340 LESAs is within the 95% confidence interval of the total U.S. LESA estimate reported by the CESS study.

Conclusion

The analysis that has been reported here is rather simple and superficial. Some of the assumptions that have been made border on being outrageous. Neverthess, the results of the analysis indicate that these assumptions do not create meaningful differences in the final estimates.

C: Classification Errors in Selection of a Criterion Score on the Language Measurement and Assessment Inventory (Table A-4)

REVISED

TABLE A-4

Classification Errors in Selection of a Criterion Score on the Language Measurement and Assessment Inventory

AGE	Proficiency in English on the Predictor	Proficiency in English on the Criterion[a] Fluent	Limited	Critical Score	Percent Accuracy[b]
5	Fluent	32	0	19	90.0
	Limited	7	30		
6	Fluent	27	0	26	87.3
	Limited	9	35		
7	Fluent	31	0	39	89.6
	Limited	7	29		
8	Fluent	36	0	39	95.3
	Limited	3	25		
9	Fluent	35	0	43	91.0
	Limited	6	26		
10	Fluent	35	0	49	91.5
	Limited	6	30		
11	Fluent	34	0	41	82.1
	Limited	14	30		
12	Fluent	27	6	47	83.6
	Limited	3	19		
13	Fluent	42	0	48	92.5
	Limited	5	20		
14	Fluent	39	10	52	82.9
	Limited	4	29		

a. Entries are number of cases in field test three.

b. For example, percent correct at age 5 equals 100 (32+30)/69=90.0.

D: Values of L, F, P_{11}, P_{12}, P_{21}, P_{22}, N_1 and N_2 for the Minimized Misclassifications of LEP Children by Age Cohort

Appendix D. Values of L, F, P_{11}, P_{12}, P_{21}, P_{22}, N_1 and N_2 for the Minimized Misclassifications of LEP Children by Age Cohort

Age Cohort	L	F	P_{11}	P_{12}	P_{21}	P_{22}	N_1	N_2
5	254657	73213	1.000	0.063	0.000	0.937	249734	78136
6	303584	90989	0.963	0.091	0.037	0.909	306970	87603
7	318470	144466	0.968	0.056	0.032	0.944	320774	142162
8	280256	37083	1.000	0.071	0.000	0.929	277422	39917
9	188187	143003	0.971	0.031	0.029	0.969	189277	141913
10	330979	32565	1.000	0.056	0.000	0.944	329047	34497
11	271485	132731	0.971	0.091	0.029	0.909	266706	137510
12	208426	191107	0.879	0.136	0.121	0.864	207388	192145
13	229986	94240	0.976	0.080	0.024	0.920	227732	96494
14	245045	241365	0.877	0.121	0.123	0.879	246282	240128

E: Contributors

CONTRIBUTORS

This report was completed while the author was assigned to the National Center for Education Statistics (NCES) as an Education Policy Fellow sponsored by the Institute for Educational Leadership, The George Washington University and Dr. Rolf M. Wulfsberg, the Assistant Administrator for Research and Analysis at NCES. The author gratefully acknowledges Dr. Wulfsberg's encouragement and support during the development and completion of this project.

Dr. Roslyn A. Korb (Technical Planning Officer, NCES) and Dr. Donald McLaughlin (Statistical Analysis Group in Education, NCES and the American Institutes for Research) provided technical opinions on the Children's English and Services Study (CESS) in the formative stage of this methodological review.

Dr. Edward A. DeAvila (Linguametrics Group) and Dr. Rudolph C. Troike (Director, National Clearinghouse for Bilingual Education) provided psychometric and linguistic perceptions on language development, measurement, and assessment.

Dr. Rebecca Oxford (InterAmerica Research Associates, Inc.) shared several references on bilingual education with the author.

Those individuals who responded to the issues paper (appendix A) provided the author with historical information on the development of the CESS which was essential for obtaining an in-depth understanding of the study.

Finally, Mr. Richard Haber (Division of Multilevel Education Statistics, NCES) completed several programming tasks for this project.